Metabolomics technologies and metabolite identification

Sofia Moco, Raoul J. Bino, Ric C.H. De Vos, Jacques Vervoort

Metabolomics studies rely on the analysis of the multitude of small molecules (metabolites) present in a biological system. Most commonly, metabolomics is heavily supported by mass spectrometry (MS) and nuclear magnetic resonance (NMR) as parallel technologies that provide an overview of the metabolome and high-power compound elucidation. Over and above large-scale analysis, a major effort is needed for unequivocal identification of metabolites. The combination of liquid chromatography (LC)-MS and NMR is a powerful methodology for identifying metabolites. Better chemical characterization of the metabolome will undoubtedly enlarge knowledge of any biological system.

Keywords: Bioinformatics; Database; Identification; Liquid chromatography; Mass spectrometry; Metabolite; Metabolome; Metabolomics; Metabonomics; Nuclear magnetic resonance

Sofia Moco*, Jacques Vervoort
Laboratory of Biochemistry, Wageningen University, Dreijenlaan 3, 6703 HA, Wageningen, The Netherlands

Sofia Moco, Raoul J. Bino, Ric C.H. De Vos
Plant Research International, P.O. Box 16, 6700 AA, Wageningen, The Netherlands
Centre for BioSystems Genomics, P.O. Box 98, 6700 AB Wageningen, The Netherlands

Raoul J. Bino
Laboratory of Plant Physiology, Wageningen University, Arboretumlaan 4, 6703 BD Wageningen, The Netherlands

*Corresponding author.
Tel.: +31 317 482620;
Fax: +31 317 484801;
E-mail: sofia.moco@wur.nl

1. Introduction

Metabolomics enables the study of the metabolic composition of an organism or biological system, so that all metabolites are described, both secondary and primary. Hence, metabolomics stands out from any other organic compound analysis in scale and chemical diversity. Perhaps the most striking feature of metabolomics lies in its integrative capacity, as one of the 'omics' disciplines, which has resulted in a shift from mainly pure (organic) chemistry-based characterization (as in phytochemistry) to a biochemical context. Metabolomics can therefore provide valuable tools, relevant in a wide range of applications (Table 1), including insight into cellular phenomena through systems-biology approaches [1,2] (e.g., in plant biochemistry, the characterization of endogenous primary and secondary metabolites is of interest for the quality of crops and their improvement, as well for the study of physiological, ecological and developmental phenomena).

In the estimated hundreds of thousands metabolites that exist in nature, there is impressive chemical variation. With a metabolomics approach, better understanding of a biological system relies on the number of participating metabolites with known identities. The need to enlarge the list of identified metabolites is a major constraint in metabolomics studies (Fig. 1). The extensive datasets obtained nowadays from analytical platforms (e.g., the most commonly used mass spectrometry (MS)-based and NMR-based systems) create a gap between ''signal x (or at most, detected metabolite X)'' and ''metabolite with IUPAC name 2-(3,4-dihydroxyphenyl)-5,7-dihydroxy-3-[(2S,3R,4S,5R,6R)-3, 4, 5-trihydroxy-6-[[(2R,3R,4R,5S,6S)-3,4, 5-trihydroxy-6-methyl-oxan-2-yl]oxymethyl]oxan-2-yl]oxy-chromen-4-one, also commonly known as rutin with CAS registry number 153-18-4, described using a certain InChi identifier'' (Fig. 2). In fact, the assignment of metabolites from experimental data (e.g., photodiode array (PDA) and mass chromatograms, and MS and NMR spectra) is a challenging task that profits from the integration of analytical tools. The combination of LC-MS with NMR is a powerful strategy in assignment and elucidation of structure to metabolites from complex extracts.

Emerging developments in analytical technologies can provide more information from the experimental data generated, leading to assignment of metabolites:

- fast, high-resolution separation systems (e.g., ultra-performance liquid chromatography (UPLC));
- high-mass accuracy and large dynamic-range MS instruments that allow extraction of reliable accurate mass values and isotopic distributions for molecular formulae (MF) calculation; and,

0165-9936/$ - see front matter doi:10.1016/j.trac.2007.08.003

Table 1. Fields of application of metabolomics

Plant breeding and assessment of crop quality
Food assessment and safety
Toxicity assessment
Nutrition assessment
Medical diagnosis and assessment of disease status
Pharmaceutical drug development
Yield improvement in crops and fermentation
Biomarker discovery
Technological advances in analytical chemistry
Genotyping
Environmental adaptations
Gene-function elucidation
Integrated systems biology

- higher sensitivity NMR systems with possibilities for on-line MS hyphenation.

Furthermore, identification of metabolites is a challenge that resides in not only obtaining high-quality data suitable for identification, from the available analytical technologies, but also integration and development of bio-computational tools for automation of data analysis. Construction of (experimental) spectrometry-based and spectroscopy-based metabolite databases and accessibility to searchable chemical databases are some of the initiatives that can aid narrowing the gap between spectrometric or spectroscopic signals and the metabolite (identified).

Analytical and computational technologies used in metabolomics allow the characterization of molecules by providing data that can lead to annotation and ultimately to identification.

In this study, we pinpoint several major considerations to be taken into account in any metabolomics approach:

- sample preparation;
- analytical technique used;
- data analyses;
- identification tools and databases; and, finally
- hypothesis testing and conclusions.

We pay special attention to identifying metabolites in plants using LC-MS and NMR strategies.

2. Sample preparation

Sample preparation is perhaps the most underestimated part of metabolomics analyses. In any biological system, metabolites of a wide chemical diversity are present in a dynamic range of concentrations that can exceed 10^6 (e.g., ratio of concentrations between sucrose and brassinolide).

In plants, a major part of the large diversity in the metabolome is due to the presence of a wide range of secondary metabolites, which generally greatly exceeds the number of primary metabolites. The composition and the quantity of metabolites detected depend to a large extent on the sample preparation chosen. The large chemical variation in plants exists between not only different plant species but also different tissues of a single plant. According to Krishnan et al. [3], a typical cell may contain 5000 metabolites (expected to be in diverse concentrations and with diverse chemical properties), which challenge the ability of a sample-preparation method to capture as many of these metabolites as possible. The extent of the detected metabolome therefore depends on the contents of the (prepared) biological sample. The more steps in sample preparation (e.g., sequential extractions and concentrations to favor a particular class of compounds), the narrower will be the chemical diversity of compounds present in the final extract. One should be aware that the further into the analysis pipeline, the slenderer will be the overall

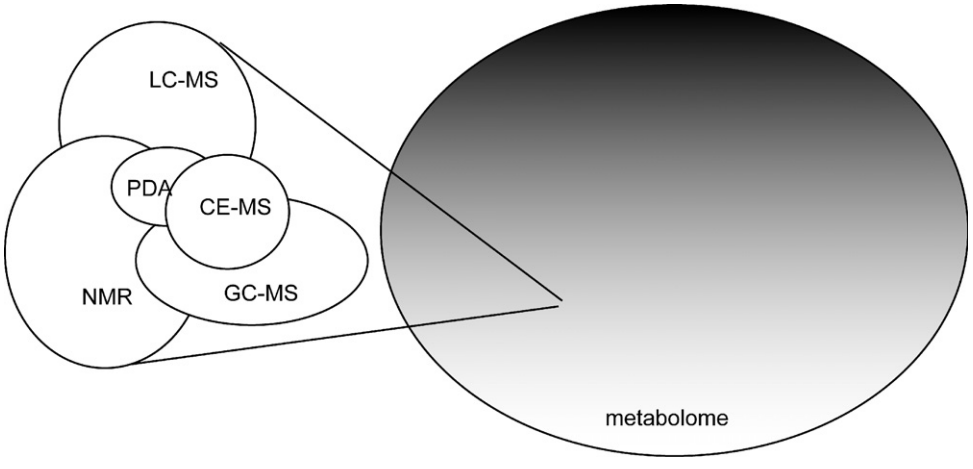

Figure 1. Heuristic representation of the metabolome, indicating that only a small fraction of metabolites have been identified so far, the majority of naturally-occurring metabolites still being unknown. The most commonly used analytical techniques that have been used for metabolite identification are LC-MS, NMR, GC-MS, CE-MS and PDA.

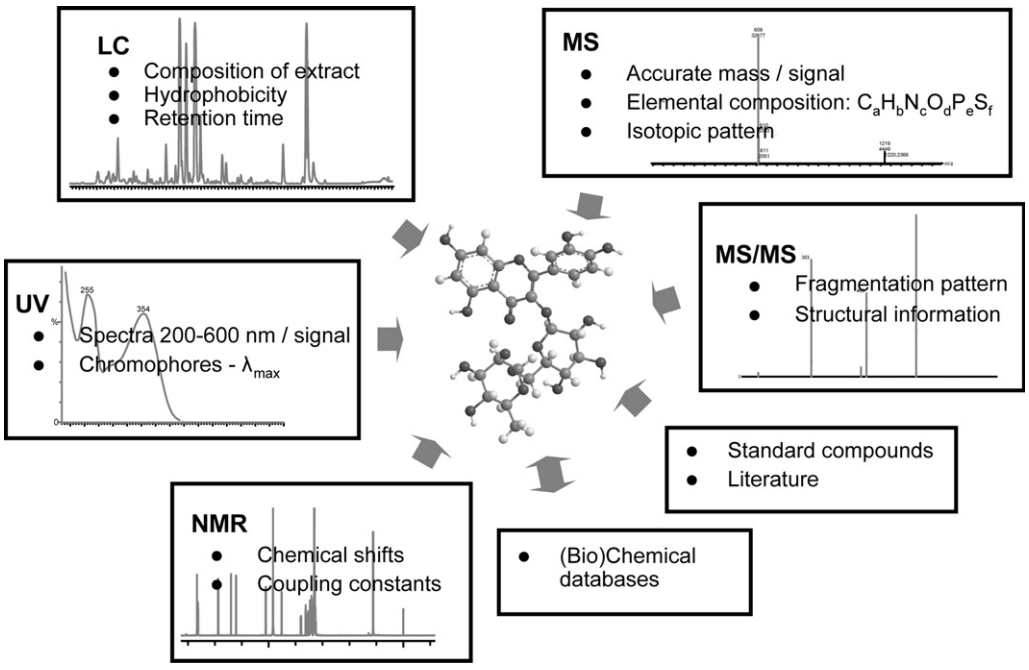

Figure 2. Data given by analytical technologies and databases that can lead to identification of a metabolite for the example of rutin (IUPAC name: 2-(3,4-dihydroxyphenyl)-5,7-dihydroxy-3-[(2S,3R,4S,5R,6R)-3,4,5-trihydroxy-6-[[(2R,3R,4R,5S,6S)-3,4,5-trihydroxy-6-methyl-oxan-2- yl]-oxymethyl]oxan-2-yl]oxy-chromen-4-one; CAS registry number: 153-18-4; InChi identifier: 1/C27H30O16/c1-8-17(32)20(35)22(37)26(40-8)39-7-15-18(33)21(36)23(38)27(42-15)43-25-19(34)16-13(31)5-10(28)6-14(16)41-24(25)9-2-3-11(29)12(30)4-9/h2-6,8,15,17-18,20-23,26-33,35-38H,7H2, 1H3/t8-,15+,17-,18+,20+,21-,22+,23+,26+,27-/m0/s1): liquid chromatography (LC); mass spectrometry (MS); fragmentation pattern analysis (MS^2); ultraviolet/visible spectroscopy (UV/Vis); nuclear magnetic resonance (NMR) experimental validation by standard compounds; and, information present in literature and databases.

knowledge about the metabolites present in the sample. However, knowledge about the (narrow) set of metabolites that survives the complete analytical pipeline (i.e. from sample preparation to identification) is progressively richer (Fig. 3).

In order to have reproducible measurements, the conditions of the biological material should be as homogeneous as possible, in terms of environment (e.g., light, temperature, humidity, nutrients, time of sampling), ideally leaving the biological variation as the only inherited variation. For metabolomics applications, a fast, reproducible, unselective extraction method is preferred for detecting the wide range of metabolites that occur in a plant, avoiding unforeseen chemical modifications. There are various methodologies for extracting compounds from biological materials:

- liquid extraction (temperature- or pressure-assisted);
- solid-phase extraction (SPE);
- solid-phase microextraction (SPME); and,
- microwave-assisted extraction (MAE).

In general, metabolites of interest are extracted by liquid extraction with one solvent, aqueous or organic, or with a combination of solvents (liquid-liquid extraction), implying that the type of metabolites extracted depends on the chemical properties of the solvent used. For a certain class of metabolites, a particular solvent can be more adequate, yet not unique for its extraction. For plants, semi-polar compounds (e.g., phenolic acids, flavonoids, alkaloids and glycosylated sterols) are successfully extracted in solutions of methanol/water, while the apolar carotenoids are better extracted in chloroform. The choice of solvent should also be compatible with the analytical instruments used. For reversed phase LC-MS analyses, solvents such as ethyl acetate or chloroform are not advisable, as these do not dissolve in the mobile phase used for the chromatography nor do they produce an efficient spray in the case of flow-injection analysis. However, in NMR analyses, any solvent can be used, preferentially deuterated in case of ^{1}H-NMR measurements. More important than the choice of sample-preparation protocol is the reproducibility of the extraction and the ability to distinguish naturally-occurring compounds.

3. LC-MS

Most MS applications in metabolomics use a separation method before mass detection, typically LC, gas chromatography (GC) or capillary electrophoresis (CE). GC-MS and LC-MS are widely used techniques and can detect a wide variety of compounds. However, the

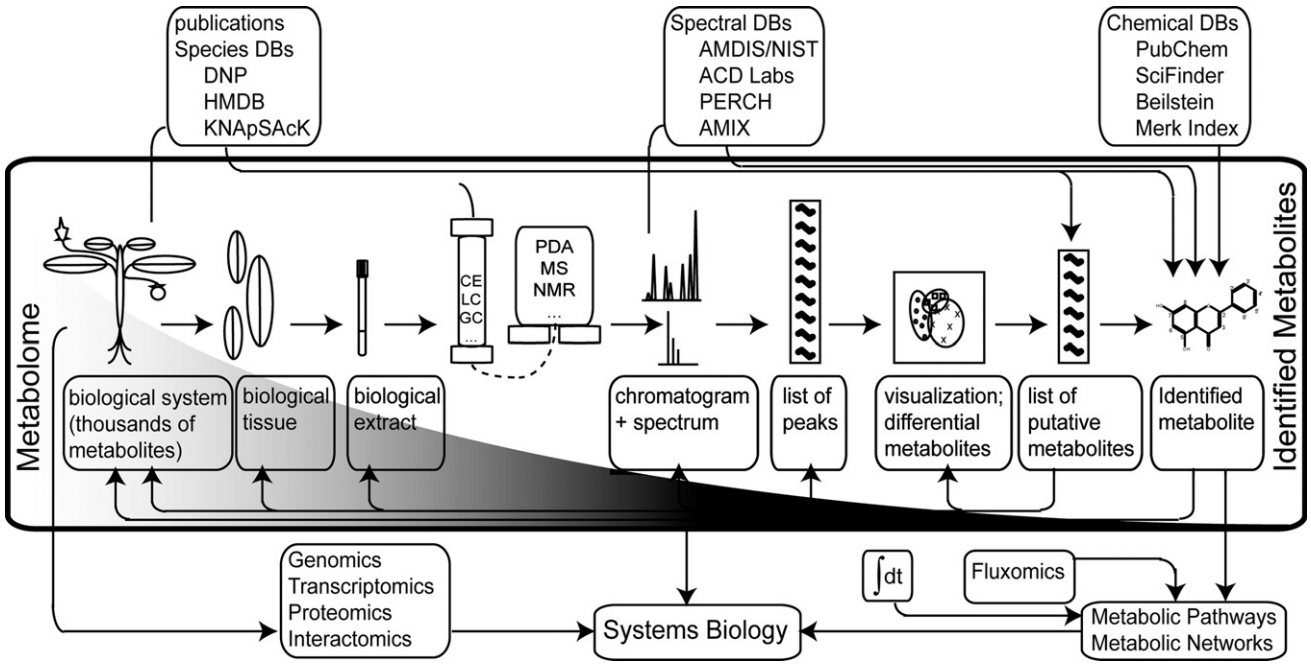

Figure 3. Metabolomics pipeline towards a systems biology approach: from the whole metabolome to identified metabolites. The large number of metabolites present in a biological system (e.g., plant) requires a strategy combining experimental design and data interpretation to identify only a few metabolites naturally occurring in the system. This procedure includes: sampling; sample preparation and extraction; analysis by, typically, CE, LC, GC or NMR; interpretation of chromatograms and spectra obtained; listing statistically relevant candidate peaks; visualization of the multivariate data; extraction of differential peaks; construction of a list of putative metabolites; and, identification of a metabolite. During this procedure, resources (e.g., species databases, literature, spectral databases and chemical databases) can be fed into the analytical pipeline to narrow the number of ambiguities for candidate metabolites. Metabolite information can aid the interpretation of metabolic pathways and networks in combination with dynamic and transient measurements made by flux analyses. The integration of genomics, transcriptomics, proteomics and interactomics with metabolomics contributes to a systems biology overview of the system (for abbreviations see Table 2).

configuration of MS instruments for these two methods is distinct due to the ionization procedures used; GC-MS instruments make use of the hard-ionization method, electron-impact (EI) ionization, while LC-MS mostly uses soft-ionization sources (e.g., atmospheric pressure ionization (API) (e.g., electrospray ionization (ESI)) and atmospheric pressure chemical ionization (APCI)) (Fig. 4).

LC is probably the most versatile separation method, as it allows separation of compounds of a wide range of polarity with little effort in sample preparation (compared to GC-MS). Using reverse-phase columns, semi-polar compounds (phenolic acids, flavonoids, glycosylated steroids, alkaloids and other glycosylated species) can be separated, and, using hydrophilic columns, polar compounds can be measured (sugars, amino sugars, amino acids, vitamins, carboxylic acids and nucleotides) [4]. UPLC can improve speed of analysis, but, more importantly, provide better chromatographic resolution in comparison to high-performance liquid chromatography (HPLC). The hyphenation of UPLC to MS can be advantageous for better assignment of metabolites from chromatographic mass signals.

MS is a spectrometric method that allows the detection of mass-to-charge species that can point to the molecular

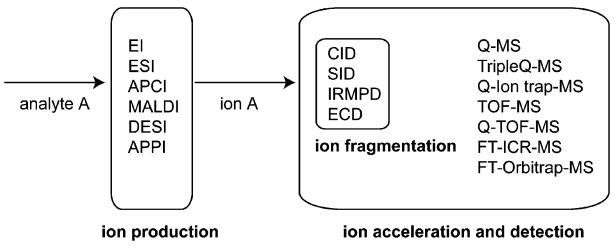

Figure 4. Possible configurations of mass spectrometers. There are different configurations of mass spectrometers, according to the ion acceleration and detection: quadrupole-MS (Q-MS); triple quadrupole-MS (TripleQ-MS); quadrupole-ion trap-MS (Q-Ion trap-MS); time-of-flight-MS (TOF-MS); Fourier transform-ion cyclotron resonance-MS (FT-ICR-MS); and, FT-Orbitrap-MS. In addition, there are different interfaces for the production of ions: electron impact (EI); electrospray (ESI); atmospheric pressure chemical ionization (APCI); matrix-assisted laser desorption ionization (MALDI); desorption electrospray ionization (DESI); and, atmospheric pressure photoionization (APPI). In terms of ion-fragmentation techniques, collision-induced dissociation (CID) is the most conventional method. Other fragmentation techniques include: surface-induced dissociation (SID); and, infrared multiphoton dissociation (IRMPD). Recently, there has been interest in electron-capture dissociation (ECD), especially for fragmentation of multiply-charged polypeptides.

mass (MM) of the detected metabolite. As MS is a developing technology in metabolomics applications, there are various configurations of mass spectrometers used for LC-MS applications, in terms of ion acceleration and mass detection, ion-production interfaces and ion-fragmentation capabilities (Fig. 4). Moreover, over the years, there have been constant adjustments in the hardware and the software of mass spectrometers to meet the demands for robustness, practicality, applicability and efficiency of the analyses.

The performance of soft-ionization mass spectrometers, used in LC-MS applications, can be described (and compared) by means of several intrinsic parameters (Fig. 5):

- mass resolving power (or resolution);
- mass accuracy;
- linear dynamic range; and,
- sensitivity [5].

Improvement of these parameters enables more effective identification of the MM of the analyte injected into the MS instrument. In general, the much used quadrupole (Q)-MS instruments have a mass resolving power that is 4 times less than that of a time-of-flight (TOF)-MS, while a Fourier transform (FT)-ion cyclotron (ICR)-MS can reach a resolving power of higher than 1,000,000 (i.e. 400 times greater than a Q-MS) [6]. A higher mass accuracy facilitates a finer distinction between closely-related mass-to-charge signals, so the quality and the quantity of assignments of mass signals to metabolites can be much improved by using high-resolution and ultra-high-resolution accurate mass spectrometers.

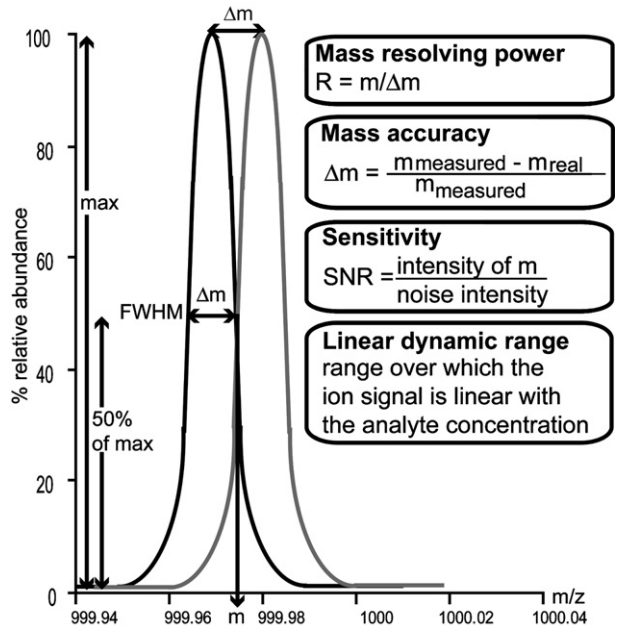

Figure 5. Parameters used to describe the performance of mass spectrometers. The mass-resolving power (or resolution), $m/\Delta m_x$, can be described in two ways: (i) m being the averaged mass-to-charge ratio associated with two adjacent mass signals of equal size and shape that overlap by x% (50% is commonly used nowadays) and Δm_x being the difference in mass-to-charge between the two adjacent mass signals; or, (ii) m being the mass at the apex of the mass signal and Δm_x being the width at x% height (typically 50%) of this mass signal, designated by FWHM (full width at half height of maximum). The mass accuracy is described by the ratio between the mass error (difference between measured and real mass) and the theoretical mass, often represented as parts per million (ppm). The sensitivity is described by the ratio between the intensity level of the mass signal and the intensity level of the noise. The linear dynamic range is described as the range of linearity of the ion signal measured as a function of the analyte concentration.

Hybrid TOF-MS instruments (e.g., Q-TOF-MS) are widely used in metabolomics due to their high-sensitivity mass-resolving power (about 10,000) and mass accuracy, having a semi-automated instrument control. However, in terms of linear dynamic range, (Q)TOF-MS instruments are limited by the properties of the time-to-digital converter detector that is only able to record one ion per dead time. Intense mass signals become saturated, masking their real intensity and leading to distortions on the mass-peak shape, producing deviations in the mass accuracy, typically to lower mass-to-charge values [7,8]. Recently, some improvements to (Q)TOF-MS instruments have been implemented, extending their dynamic range. The use of an on-line lock mass spray, acting as an internal standard, can help to correct for deviations in the mass-to-charge axis, and can dictate the ion-intensity interval for which the mass accuracy is highest and adequate for calculating elemental composition [9].

FT-MS instruments, both the cyclotron (FT-ICR-MS) and the Orbitrap type (FT-Orbitrap-MS), enable measurements at a higher mass accuracy in a wider dynamic range. FT-ICR-MS has the highest mass-resolving power so far reported for any mass spectrometer (>1,000,000) and a mass accuracy generally within 1 ppm [10]. The recently developed FT-Orbitrap-MS has a more modest performance compared to the FT-ICR-MS (maximum resolving power >100,000 and 2 ppm of mass accuracy with internal standard), but it is a high-speed, high-ion-transmission instrument, due to shorter accumulation times. This is a very advantageous characteristic, especially when hyphenated to a separation technique, such as LC, and also when carrying out MS^2 experiments [11]. The appearance of high-mass-accuracy instruments with a wide dynamic range can improve immensely identification capabilities in the on-line methods applied to complex mixtures.

The mass detection of a molecule by MS, in LC-MS applications, is conditioned by the capacity of the analyte to ionize while being part of a complex mixture. Because only ions, either anions or cations, can be measured by MS, metabolites unable to ionize cannot be detected. Apart from the chemical properties of the molecule itself, eluent flow and composition, sample matrix and ionization source all influence ionization. Ion-suppression and matrix effects can become major

issues, particularly in semi-quantitative measurements. The use of ionization enhancers, sample-clean-up methods and different ionization sources are some of the possibilities that can improve ionization of analytes under study [12].

Regardless of the configuration of the (UP)LC-MS system, robustness and reproducibility (in retention time and mass accuracy) as well as efficient ionization of the analyses are essential for obtaining consistent data [13]. The chromatographic parameters (temperature, pH, column, flow rate, eluents, gradient), injection parameters, sample properties, MS and MS^2 parameters (calibration and instrumental parameters, such as capillary voltage and lens orientation), and all other parameters related to the configuration of the LC-MS system (e.g., presence of other detectors, such as PDA, and tube widths) may all influence the performance of metabolomics analyses. An adequate configuration should be adopted, fitting to the aim of the analyses and the limitations of the instruments.

3.1. Identifying metabolites

Metabolite assignments using LC-MS as a tool for compound identification are usually obtained by combining accurate mass, isotopic distribution, fragmentation patterns and any other MS information available.

Calculation of the chemical combinations that fit a certain accurate mass is generally one of the first steps to obtain a set of alternatives that can lead to the identity of the metabolite detected. This set of alternatives becomes less extensive if the mass spectrometer can provide a more accurate MM value [14]. Using an instrument that can provide very high mass accuracies, the range of possibilities of MFs is limited and can, especially for lower m/z values, lead to the correct MF. The number of possible MFs increases with increasing MM values. Furthermore, in most cases, chemical elements can be pre-selected, avoiding generation of excessive false alternatives from including all elements of the periodic table. For general applications in plant or animal metabolomics, most metals can be excluded (except perhaps for Na or K that are common adducts in mass spectra), the core elements being C, H, O, N, P and S. Logically, any other element for which there is the slightest evidence of being present in the analyzed sample should be included in calculating elemental composition.

Another aspect to take into account when MFs are calculated from MMs is the algorithm used for the calculation. There are more possible mathematical combinations of elements that fit certain MMs than the number of MFs that exist chemically. This is related to chemical rules (e.g., the octet rule) that dictate certain limitations on chemical bonding derived from the electronic distribution of the participating atoms present in molecules. The widely-applied nitrogen rule is used for assessing the presence or absence of N atoms in a molecule or ion. Another useful factor is the presence of rings and double bonds. As described by Bristow [15], the number of rings and double bonds can be calculated from the number of C, H and N atoms that a molecule contains.

One of the most powerful methods for narrowing the number of MFs is to make use of the isotopic pattern of a mass signal. For most small organic molecules, M, the intensity of the second isotopic signal, corresponding to the ^{13}C signal, can indicate the number of carbons that the molecular ion contains using the knowledge that the natural abundance of ^{13}C is 1.11%, so this is of major assistance in assigning MFs from MMs. According to Kind and Fiehn [14], this strategy can remove more than 95% of false positives and can even outperform an analysis of accurate mass alone using a (as yet non-existent) mass spectrometer capable of 0.1 ppm mass accuracy. With the appearance of MS instruments with large dynamic range (with good isotopic intensity measurements), this is certainly an efficient strategy when combined with the MS spectra-analysis tools described below.

The fragmentation pattern of a mass signal can provide structural information about the fragmented ion. From the fragments obtained, the structure of the molecule can be deduced, knowing that the breakages will occur at the weakest points of the ion. For example, an *O*-glycosylated flavonoid will first fragment on the glycosidic linkage and only afterwards in the aglycone backbone, if sufficient energy is provided. The possibility of isolating one ion and performing MS^2 to the successively obtained fragments can be highly informative for tracking functional groups and connectivity of fragments for elucidating the structures of metabolites. In addition, the possibility of obtaining accurate mass fragments is another advantage when there is little knowledge about the possible atomic arrangements of the molecular ion.

Moreover, there is a series of possible MS experimental procedures that can enhance our knowledge about the metabolites of interest. These experiments include comparisons of analyses obtained by positive and negative ion modes (either by on-line switching or off-line), neutral mass-loss experiments that can aid identification of certain functional groups or substituents (e.g., hydroxyls, carbonyls or glycosides [16]).

The usage of ^{13}C material as internal standard is also an elegant method of obtaining metabolite information [17].

In addition, when a separation method is coupled to the mass spectrometer, retention time is a parameter that can give information about the polarity of the metabolite. Nowadays, in stabilized (LC or GC)-MS setups, retention-time variation can be relatively low, allowing direct comparisons of chromatograms and the construction of metabolite databases [9,13,18].

Data obtained from additional detectors can also be a complementary source of structural information about a metabolite. Typically, for a well-separated

chromatographic signal with sufficient intensity, a full absorbance spectrum can be obtained in the ultraviolet/visible (UV/Vis) range using a PDA detector. For many secondary metabolites, their light-absorbance spectra can indicate at least the classes of compounds to which these belong, as the type of chromophores can be inferred from the absorbance maxima and the shape of the spectrum (e.g., due to their polyaromatic system giving specific absorbance maxima that undergo slight shifts with the introduction of conjugations in the polyaromatic system).

Possibly the most straightforward approach to obtaining confirmation of the identity of metabolites in a biological sample is to test commercially available standard compounds on the same analytical system. However, this approach implies (commercial) availability of such standard compounds, which are scarce, especially in the secondary metabolism field. In addition, many standard compounds used in metabolomics are unstable and/or impure. This can especially be problematic in MS if the ionization efficiency of the impurities is relatively high. Nevertheless, when standard compounds are available, these are useful for not only confirmation of the identity of compounds but also undergoing (semi-)quantitative analyses and, most importantly, construction of metabolite databases containing experimental data of tested compounds on a fully characterized system.

In summary, the ability to assign metabolites using MS resides in the possibility of combining different features of the MS analysis (accurate mass, fragmentation pattern, and isotopic pattern) with additional experimental parameters (e.g., retention time, and UV/Vis spectra) and confirmation with standard compounds. Also, biochemical, literature and species information, and related relevant information is taken into account in assigning the metabolite being studied (Fig. 2).

4. NMR

NMR is a spectroscopic technique that takes advantage of the spin properties of the nucleus of atoms. Because the nuclear transition energy is much lower (typically of the order of 10^4) than an electronic transition, NMR is not as sensitive as other techniques, such as infrared (IR) or UV/Vis spectroscopy [19,20].

The signal-to-noise ratio (S/N) in NMR depends on many parameters (Fig. 6) (e.g., magnetic field strength of the instrument (B_0), concentration of the sample, acquisition time (NS), and the measurement temperature). A lower temperature can increase the S/N by influencing the Boltzmann equilibrium, but the temperature also influences the T_2^* of the signals measured. The T_2^* is inversely related to the line-width ($\Delta\nu 1/2$) of the signals obtained ($\pi\Delta\nu 1/2 = 1/T_2^*$) and is influenced by

$$\frac{S}{N} \propto NAT^{-1}B_0^{3/2}\gamma_{exc}\gamma_{obs}^{3/2}T_2^*(NS)^{1/2}$$

S/N = signal-to-noise ratio
N = number of molecules in the observed sample volume
A = abundance of the NMR active spins involved in the experiment
T = temperature
B_0 = static magnetic field
γ_{exc} = magnetogyric ratio of the initially excited spins
γ_{obs} = magnetogyric ratio of the observed spins
T_2^* = effective transverse relaxation time
NS = total number of accumulated scans

Figure 6. Signal-to-noise ratio (S/N) equation in NMR spectroscopy [19].

the tumbling rate of the molecule and also inhomogeneities in the magnetic field. Usually, for small molecules, faster tumbling rates (often linked to higher measurement temperatures) provide narrower line-widths. Magnetic field inhomogeneities can be caused by magnetic field-susceptibility fluctuations in the sample (e.g., presence of large particles, paramagnetic ions or inferior NMR tubes) or by poor shimming. Automated shimming procedures in modern NMR instruments largely alleviate poor shimming, leaving sample preparation as the major cause of inferior NMR spectra.

NMR is perhaps the most selective analytical technique available, being able to provide unambiguous information about a molecule. NMR can elucidate chemical structures, and can provide highly specific evidence for the identification of a molecule. Furthermore, NMR is a quantitative technique, as the number of nuclear spins is directly related to the intensity of the signal [21].

Different metabolomics approaches can be applied when using NMR [22]. One of the approaches is directly related to the usage of NMR for structure elucidation. ^{1}H is the most used nucleus for NMR measurements due to its very high natural abundance (99.9816–99.9974% [23]) and good NMR properties. In general, the compounds of interest are isolated from their tissues, often through laborious analytical procedures, and solubilized for the acquisition of one-dimensional ^{1}H NMR and, when required, additional two-dimensional (2D)-NMR spectra. For most bio-organic compounds, the acquisition of a 1D ^{1}H NMR spectrum is not sufficient for full structural elucidation. Homonuclear ^{1}H-2D spectra (e.g., COSY [24], TOCSY [24] and NOESY [24]) and heteronuclear 2D spectra can be acquired for detecting direct ^{1}H-^{13}C bonds by HMQC [24] or, over a longer range, HMBC [24]. Heteronuclear 2D-NMR data acquisition is more demanding on instrument time than homonuclear 2D-NMR data acquisition but the information from heteronuclear 2D spectra is extremely useful for identifying unknown molecules [24].

A fast-growing approach, particularly in animal and human research, is NMR fingerprinting, which involves acquiring NMR spectra of complex mixtures, as biofluids

or plant extracts, to pinpoint differences between the samples, with the goal of biomarker discovery [22,25]. Fingerprinting with both NMR and MS gives a global overview of the metabolome. Tomato fruits and Arabidopsis leaves have been profiled by NMR [26,27]. Most studies so far use ^{1}H NMR, as being the least selective for the type of molecules, and that can provide the highest sensitivity. However, ^{13}C NMR [28] and 2D measurements (e.g., JResolved (JRes) [29], COSY [30] and HMBC [31]) have also been used for profiling.

In NMR profiling, the necessity of spectral comparisons demands that spectrum acquisition and control of conditions should be extremely rigorous. Small changes in temperature, pH, and presence of impurities or degradation of the sample material can lead to detection of false metabolic changes and hence incorrectly indicate different metabolites.

Nowadays, in a 14.1 Tesla (600 MHz for ^{1}H NMR) instrument, the limit of detection is in the microgram (^{1}H-^{13}C NMR) or even sub-microgram region (^{1}H NMR). The sensitivity of NMR has been improving over the years, increasing the suitability of this technique for analytical applications. Labeling low-abundant metabolites with stable isotopes (^{13}C/^{15}N) can be applied as a strategy for performing 2D-NMR analysis on low amounts of material. In flux analysis, compounds are labeled for analysis of the propagation of the isotope label in pathway analysis and kinetics measurements [32].

The appearance of cryogenic probe-heads brought important improvements in NMR sensitivity [33]. Being able to take advantage of the reduction in thermal noise by using low-temperature detection coils, S/N can be obtained up to five-fold higher than by using conventional probes. In addition, the possibility of miniaturizing the active volume of the detection cell enabled the appearance of microprobes. Moreover, the S/N of the detection coil is inversely related to its diameter. These miniaturized NMR probes are available with active volumes as low as 1.5 μL, providing new possibilities for analyzing molecules in the lower detection volumes, increasing the concentration of the analyte at no expense to the S/N. This low active volume is compatible with chromatographic elution volumes in capillary chromatography, making capillary microcoil NMR (CapNMR) feasible [34].

4.1. LC-(SPE)-NMR

The coupling of LC with NMR is becoming increasingly useful as NMR sensitivity improves, avoiding excessive analytical demands on obtaining enough material to perform NMR measurements. In practical terms, LC-NMR is still not as clear-cut as LC-MS but is establishing itself as a powerful system for identifying related metabolites from complex mixtures (e.g., natural extracts from plants).

There are different configurations for coupling LC to NMR [35]. More recently the on-line coupling of LC to SPE and subsequent NMR became available and improved some of the existing analytical barriers of the previous modes. In this configuration, the chromatographic peaks are trapped in SPE cartridges and can be concentrated up to several times by multi-trapping into the same cartridge. The chromatography itself can be done with (less expensive) protonated solvents because the analytes within the cartridges are dried and then eluted with fully deuterated solvents.

As an example of the efficiency of this approach, the separation of flavonoids and phenolic acids present in Greek oregano extract was accomplished by LC-SPE-NMR-MS [36]. The compounds were separated by LC, trapped in SPE cartridges, eluted for NMR and MS acquisition. Even two related flavonoids, naringenin and apigenin, co-eluting in the LC (and therefore trapped into the same cartridge) could be readily distinguished by MS and NMR [36]. This method is suitable for the analysis of less abundant compounds in complex mixtures, since it allows separation, concentration and NMR acquisition of metabolites within a single system, avoiding the often tedious analytical preparations before NMR analysis.

4.2. Identifying metabolites

The magnetic resonance of a nucleus present in a molecule is displayed as a signal with a determined frequency, represented by a chemical shift value, δ, in an NMR spectrum. The analysis of an NMR spectrum can be extremely puzzling due to overlapping signals and multiplicities within the signals. The NMR spectrum of a particular molecule is unique, and, for this reason, NMR is considered one of (and perhaps even) the most selective techniques for compound elucidation.

For the analysis of NMR spectra, the number, positions and areas of the signals in the spectrum as well as the multiplicity of these are some of the aspects that are used in order to attempt the assignment of a molecule. An aspect that can be both highly informative and difficult to interpret is the multiplicity of signals. Signal splitting or the multiplicity of the signals is caused by the spin-spin coupling between the proton and nearby atoms. The coupling constants, J_{ab}, transmit structural information, necessary for the elucidation of most molecules.

The interpretation of NMR spectra can be quite demanding, especially for highly-related structures or higher MM molecules. There are several software tools (ACD/Labs, ChemOffice, and PERCH Solutions) that can help in ^{1}H NMR spectral analysis by providing NMR spectral predictions. The aim of these prediction tools is to aid analysts to assign spectral δs and Js to the analyzed molecule. Strictly theoretical calculations of NMR spectra from molecular properties are an option, yet unaccounted effects often appear on experimental spectra and are difficult to incorporate in theoretical

prediction routines. In particular, the prediction of ^{1}H NMR spectra proves to be more difficult to implement due to the effect of 3D conformational structures on the ^{1}H NMR chemical shifts of the protons. The construction of prediction models based on experimental data can be a successful alternative in order to describe chemical phenomena at a detailed molecular level [37]. Using such an approach, identification of flavonoids from natural extracts can be feasible with the acquisition of (only) ^{1}H NMR. The compounds syringetin-3-*O*-galactoside and syringetin-3-*O*-glucoside cannot be distinguished by MS, as these have the same mass and conformation, but can be clearly distinguished by ^{1}H NMR. It is therefore possible to identify flavonoid derivatives using an NMR-based database of flavonoids [37].

4.3. LC-MS-NMR

The identification of metabolites can be aided by metabolite-profiling methods, such as MS or NMR, but often the full chemical description of a molecule is achieved only by integrating metabolite information taken from different sources. The combination of MS with NMR for unraveling the identity of a molecule is one of the most powerful strategies (Fig. 2).

MS can indicate not only the MM of a compound and therefore possible MFs, but also the presence of certain functional groups or substitution patterns. The assignment of a wide variety of metabolites from tomato extracts was feasible using LC-PDA-ESI-QTOF-MS [9]. Using this method, flavonoids, phenolic acids and alkaloids were detected from extracts of tomato fruits. Various conformational isomers, such as dicaffeoylquinic acids, have been (putatively) assigned, but not fully described chemically. In order to discern the substitution positions of functional groups and distinguish isomers, NMR is in most cases unavoidable. NMR allows the structural elucidation of molecules up to the isomer level.

In fact, the most efficient way to seize the advantages of both technologies, LC-MS and NMR, is to use them in parallel or, if possible, on-line. Coupling LC with both MS and NMR has been described and is an elegant, efficient way of obtaining useful data for identifying compounds [36]. The advantage of performing the same separation for both MS and NMR makes clear the correspondence of the chromatographic signals between these two instruments. However, due to the complex analytical set-up, it is still most common to undertake analyses by LC-MS and LC-(SPE)-NMR separately.

Developments in chemometric methods can assist the rapid identification of molecules present in complex mixtures. The method depends on data obtained from a large number of samples which are measured by both LC-MS and NMR. The different data matrices obtained from these fingerprints can be fused using concatenation or other data-fusion methods. In theory, fluctuations in the LC-MS matrix should be reflected in similar changes in the NMR matrix. When the sample preparation and analysis are done in a coherent manner, this method might enable high-throughput identification of molecules. This approach has been tested for biofluid analysis, by coupling LC-MS and NMR data of urine samples [38,39], and can be a promising strategy in biomarker discovery.

5. Data analyses

The extraction of valuable conclusions from the analysis of metabolomics data is as important as performing the analytical measurements. There are a variety of methods that allow the transformation of raw data, directly taken from the instrument, passing through different treatments and ultimately leading to a list of metabolites.

Prior to any data analysis, it is important to be aware of the possible sources of variation present in the samples that can influence the final conclusions if these are not overseen. Parameters (e.g., biological variation present among individuals, sampling, sample preparation and analytical measurement) influence reproducibility of results, and these should be monitored as much as possible by measuring replicates, both analytical and biological. In principle, biological variance should surpass all analytical variance.

Signal irreproducibility is an obstacle to reliable comparison of chromatograms and spectra. Retention-time shifts are common in GC and, more severely, LC, as are occasional shifts in NMR spectra. In NMR spectra, non-reproducibilities seem to be strictly related to sample preparation and hardly ever due to instrumental incoherence. Nevertheless, even in strictly controlled conditions, signal shifts may persist. For this reason, the use of signal-alignment software has become a routine procedure for comparing chromatograms or spectra. MetAlign [13], XCMS [40] and MZmine [41] are some of the available alignment toolboxes for MS applications, as HiRes is for NMR applications [42]. These are relevant in reducing raw data to workable datasets that are still informative.

For masking or emphasizing variable and sample deviations, scaling and standardization tools can be applied, as long as these do not lead to artificial distortions of original data. As for all the 'omics' technologies, multidimensionality is one of the characteristics of metabolomics data, which ensures that the dataset is inherently complex. Supervised and unsupervised tests (e.g., principal component analysis (PCA), hierarchical cluster analysis (HCA), partial least squares (PLS) and discriminant analysis (DA)) are widely applied in metabolomics [31,43]. These methods not only simplify the data by reducing dimensionality but also provide visual representation of the data.

Table 2. Number of metabolite records present in MS and NMR, pathway and chemical databases

Databases	Source	No. records (approximate)
MS-based		
NIST/EPA/NIH Mass Spectral Library (NIST 0.5)	National Institute of Standards and Technology (NIST)	163,000
SpecInfo	Daresbury Laboratory	139,000
Spectral Database for Organic Compounds, SDBS	National Institute of Advanced Industrial Science and Technology (AIST)	23,500
KNApSAcK (Comprehensive Species-Metabolite Relationship Database)	Nara Institute of Science and Technology (NAIST)	15,500
Metlin	The Scripps Research Institute	15,000
Human Metabolome Database (HMDB)	Genome Alberta and Genome Canada	2,300
Golm Metabolome Database (GMDB@CSB.DB)	Max Planck Institute of Molecular Plant Physiology	
Metabolome of Tomato Database (MoTo DB)	Plant Research International	100
NMR-based		
Human Metabolome Database (HMDB)	Genome Alberta and Genome Canada	400 (13C) 350 (1H)
ACD Databases	Advanced Chemistry Development, Inc.	15,000 (13C and 1H) 8,800 (15N) 26,100 (31P)
Spectral Database for Organic Compounds, SDBS	National Institute of Advanced Industrial Science and Technology (AIST)	12,500 (13C) 14,300 (1H)
SpecInfo	Daresbury Laboratory	102,000 (13C) 117,000 (1H) 1,000 (15N) 1,000 (17O) 17,000 (31P) 25,000 (19F)
Standard Compounds on Biological Magnetic Resonance Bank (BMRB)	University of Wisconsin	275 (13C and 1H)
NMRShiftDB	University of Koeln	19,500 (13C) 3,000 (1H)
Pathways		
Kyoto Encyclopedia of Genes and Genomes (KEGG)	Kyoto University / Tokyo University	14,000
Chemical		
SciFinder	Chemical Abstracts Service (CAS)	30,500,000
PubChem	National Institutes of Health (NIH)	10,100,000
Beilstein Database	MDL	9,400,000
eMolecules	eMolecules	>5,600,000
Available Chemicals Directory	Elsevier MDL	>>200,000
Combined Chemical Dictionary (CCD)	Chapman & Hall/CRC Press	
Dictionary of Organic Compounds		265,000
Dictionary of Natural Products		170,000
Dictionary of Inorganic and Organometallic Compounds		103,000
Dictionary of Drugs		44,000
Dictionary of Analytical		14,000
Reagents		
ChemIDplus	National Institute of Health (NIH)	380,000
Substance Registry System (SRS)	Environmental Protection Agency (EPA)	98,000
ChemFinder	CambridgeSoft Corporation	72,000
Merck Index	John Wiley & Sons, Inc.	10,200
Chemical Entities of Biological Interest (ChEBI)	European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI)	10,000

More sophisticated methods of emphasizing relationships between metabolites (e.g., correlation matrixes and metabolic correlation networks) can help to establish relationships between different metabolites, and even between metabolites and transcripts, genes or proteins. In this way, a systems-level overview is envisioned [44].

There are different tools for visualization or databases that can be used to display the coupling of different 'omics' data (e.g., KEGG (www.genome.jp/kegg), MetaCyc (http://metacyc.org), MAPMAN (gabi.rzpd.de/projects/MapMan) and KappaView (kpv.kazusa.or.jp/kappa-view)).

6. Identification tools and databases

There are still only a few tools that can automatically produce a list of possible metabolites from the m/z signals at a particular retention time (MS) or from chemical shifts and coupling constants (NMR). Therefore, the assignment of metabolites from experimental data implies an intensive manual effort, hindering the throughput of the analytical set-up. The bridge between experimental data (MS and NMR spectra, retention time, fragmentation pattern, chemical shift, coupling constant) and the available chemical databases (Table 2) is still weak, let alone automatic. Some identification tools (e.g., elemental composition calculation or MM calculation) exist in the software of different instruments, but these seldom allow spectral matching linked to a public database, as in proteomics applications. One of the few examples of spectral databases is AMDIS (Automated mass spectral deconvolution and identification system) (www.amdis.net), which can mostly be used for identifying GC-MS signals. Advanced Chemistry Development Labs (ACDLabs) also commercially provides spectral matching with databases for MS and NMR, as well as predictor tools. Nevertheless, many plant metabolites, such as secondary metabolites, are not present in these databases.

Building up public metabolite databases is starting to be done by laboratories within the metabolomics community (Table 2). One of the largest initiatives for the identification of metabolites is the Human Metabolome Project, which combines MS and NMR data with molecule information [45]. The detailed description of the methods of sample preparation and analysis, conditions of the analytical experiment, chemical information about the metabolites (name, IUPAC name, chemical descriptors (e.g., CAS registry numbers and InChi and/or structural information, links to chemical databases)), experimental spectra and biological source are some of the features to include in metabolite databases.

Nomenclature of molecules is a troublesome issue, as the list of common names for a given molecule can be quite extensive, and the same common name can be attributed to distinct molecules. This is a real impediment to reliable, unambiguous classification and creates false interpretations, in particular in the organization of databases and searching tools. Only with an accurate description of the experimental conditions and chemical identity of the metabolites is comparison and exchange of data relevant. Perhaps, at this stage, more priority will have to be given to rigorous identification of metabolites, as dealing with unknowns and metabolites that are not fully identified creates a lot of incongruent hits in the databases. Ideally, the separate metabolite databases will be accessible through a common search engine as an open source web service, as in BioMOBY [46].

7. Summary

The description of the metabolome can be achieved by different methods, either in parallel or in combination. MS- and NMR-profiling techniques are powerful methods for detecting the metabolome as a whole. Comparison of metabolic profiles can elucidate differences between organisms and pinpoint the metabolites responsible. However, if we can identify differences but not describe these chemically, very little is left to say about the underlying nature of the metabolic phenomena. There is still a long way to go to describe completely the metabolome of an organism, elucidation of unknowns being a priority. As yet, no single analytical method can capture the whole metabolome and the analytical method chosen defines the number of metabolites left to identify.

Currently, integration of high-resolution MS and NMR provides the necessary information for elucidation of compounds. The development of bioinformatic tools will facilitate the management of large amounts of data and help integrate different datasets by sieving the metabolite information from the instrumental chromatographs and spectra. Expansion of our view over the metabolome of organisms will improve the description of metabolic networks and cellular phenomena in general.

Acknowledgement

We acknowledge financial support from: the EU RTD project ''Capillary NMR'', a European Community-Access to Research Infrastructure action of the Improving Human Potential Program, Contract HPRI-CT-1999-00085 and Contract HPRI-CT-1999-50018; the research programme of the Centre of BioSystems Genomics (CBSG) which is a part of The Netherlands Genomics Initiative/Netherlands Organization for Scientific Research; and, the EU project ''META-PHOR'', contract number FOOD-CT-2006-036220. We would like to thank Udo Brinkman for his help with the manuscript.

References

[1] R.D. Hall, New Phytol. 169 (2006) 453.
[2] R.J. Bino, R.D. Hall, O. Fiehn, J. Kopka, K. Saito, J. Draper, B.J. Nikolau, P. Mendes, U. Roessner-Tunali, M.H. Beale, R.N. Trethewey, B.M. Lange, E.S. Wurtele, L.W. Sumner, Trends Plant Sci. 9 (2004) 418.
[3] P. Krishnan, N.J. Kruger, R.G. Ratcliffe, J. Exp. Bot. 56 (2005) 255.
[4] V.V. Tolstikov, O. Fiehn, Anal. Biochem. 301 (2002) 298.
[5] S.A. McLuckey, J.M. Wells, Chem. Rev. 101 (2001) 571.
[6] M.P. Balogh, LC-GC N. Am. 22 (2004) 118.
[7] I.V. Chernushevich, A.V. Loboda, B.A. Thomson, J. Mass Spectrom. 36 (2001) 849.
[8] H.A. Verhoeven, C.H. de Vos, R.J. Bino, R.D. Hall, in: K. Saito, R.A. Dixon, L. Willmitzer (Editors), Plant Metabolomics - Biotechnology and Forestry, Springer-Verlag, Berlin, Germany, 2006.
[9] S. Moco, R.J. Bino, O. Vorst, H.A. Verhoeven, J. de Groot, T.A. van Beek, J. Vervoort, R.C.H. De Vos, Plant Physiol. 141 (2006) 1205.
[10] S.C. Brown, G. Kruppa, J.L. Dasseux, Mass Spectrom. Rev. 24 (2005) 223.
[11] A. Makarov, E. Denisov, A. Kholomeev, W. Baischun, O. Lange, K. Strupat, S. Horning, Anal. Chem. 78 (2006) 2113.
[12] C.R. Mallet, Z.L. Lu, J.R. Mazzeo, Rapid Commun. Mass Spectrom. 18 (2004) 49.
[13] R.C.H. De Vos, S. Moco, A. Lommen, J.J.B. Keurentjes, R.J. Bino, R.D. Hall, Nat. Protoc. 2 (2007) 778.
[14] T. Kind, O. Fiehn, BMC Bioinformatics 7 (2006) 234.
[15] A.W.T. Bristow, Mass Spectrom. Rev. 25 (2006) 99.
[16] N. Fabre, I. Rustan, E. de Hoffmann, J. Quetin-Leclercq, J. Am. Soc. Mass Spectrom. 12 (2001) 707.
[17] M.R. Mashego, L. Wu, J.C. Van Dam, C. Ras, J.L. Vinke, W.A. Van Winden, W.M. Van Gulik, J.J. Heijnen, Biotechnol. Bioeng. 85 (2004) 620.
[18] J. Lisec, N. Schauer, J. Kopka, L. Willmitzer, A.R. Fernie, Nat. Protoc. 1 (2006) 387.
[19] T. Claridge, High-resolution NMR Techniques in Organic Chemistry, Elsevier, Amsterdam, The Netherlands, 1999.
[20] C.R. Nave, HyperPhysics, 2005 (http://hyperphysics.phy-astr.gsu.edu/hbase/hph.html#hph).
[21] G.F. Pauli, Phytochem. Anal. 12 (2001) 28.
[22] R.G. Ratcliffe, Y. Shachar-Hill, Biol. Rev. 80 (2005) 27.
[23] J. de Laeter, J.R., Pure Appl. Chem. 75 (2003) 683.
[24] W.F. Reynolds, R.G. Enriquez, J. Nat. Prod. 65 (2002) 221.
[25] S. Kochhar, D.M. Jacobs, Z. Ramadan, F. Berruex, A. Fuerhoz, L.B. Fay, Anal. Biochem. 352 (2006) 274.
[26] G. Le Gall, I.J. Colquhoun, A.L. Davis, G.J. Collins, M.E. Verhoeyen, J. Agric. Food Chem. 51 (2003) 2447.
[27] J.L. Ward, C. Harris, J. Lewis, M.H. Beale, Phytochemistry 62 (2003) 949.
[28] G. Vlahov, Anal. Chim. Acta 577 (2006) 281.
[29] M.R. Viant, Biochem. Biophys. Res. Commun. 310 (2003) 943.
[30] Y. Xi, J.S. de Ropp, M.R. Viant, D.L. Woodruff, P. Yu, Metabolomics V2 (2006) 221.
[31] S. Masoum, D.J.R. Bouveresse, J. Vercauteren, M. Jalali-Heravi, D.N. Rutledge, Anal. Chim. Acta 558 (2006) 144.
[32] R.G. Ratcliffe, Y. Shachar-Hill, Plant J. 45 (2006) 490.
[33] H. Kovacs, D. Moskau, M. Spraul, Prog. Nucl. Magn. Reson. Spectrosc. 46 (2005) 131.
[34] F.C. Schroeder, M. Gronquist, Angew. Chem., Int. Ed. Engl. 45 (2006) 7122.
[35] V. Exarchou, M. Krucker, T.A. van Beek, J. Vervoort, I.P. Gerothanassis, K. Albert, Magn. Reson. Chem. 43 (2005) 681.
[36] V. Exarchou, M. Godejohann, T.A. van Beek, I.P. Gerothanassis, J. Vervoort, Anal. Chem. 75 (2003) 6288.
[37] S. Moco, L.H. Tseng, M. Spraul, Z. Chen, J. Vervoort, Chromatographia 9/10 (2006) 503.
[38] D.J. Crockford, E. Holmes, J.C. Lindon, R.S. Plumb, S. Zirah, S.J. Bruce, P. Rainville, C.L. Stumpf, J.K. Nicholson, Anal. Chem. 78 (2006) 363.
[39] J. Forshed, H. Idborg, S.P. Jacobsson, Chemometr. Intellig. Lab. Syst. 85 (2007) 102.
[40] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, Anal. Chem. 78 (2006) 779.
[41] M. Katajamaa, J. Miettinen, M. Oresic, Bioinformatics 22 (2006) 634.
[42] Q. Zhao, R. Stoyanova, S.Y. Du, P. Sajda, T.R. Brown, Bioinformatics 22 (2006) 2562.
[43] M. Scholz, F. Kaplan, C.L. Guy, J. Kopka, J. Selbig, Bioinformatics 21 (2005) 3887.
[44] A.R. Joyce, B.O. Palsson, Nat. Rev. Mol. Cell Biol. 7 (2006) 198.
[45] D.S. Wishart, D. Tzur, C. Knox, R. Eisner, A.C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.-A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D.D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G.E. Duggan, G.D. MacInnis, A.M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B.D. Sykes, H.J. Vogel, L. Querengesser, Nucl. Acids Res. 35 (2007) D521.
[46] M. Wilkinson, H. Schoof, R. Ernst, D. Haase, Plant Physiol. 138 (2005) 4.

Sofia Moco obtained a Chemical Engineering degree from the Instituto Superior Técnico, Technical University of Lisbon, Portugal. She is finishing her PhD on MS and NMR-based metabolomics of plants at the Biochemistry Laboratory, University of Wageningen, and Plant Research International, Wageningen, The Netherlands.

Raoul J. Bino has studied Biology at the University of Amsterdam. He obtained his PhD and later a professorship on plant metabolomics at the Plant Physiology Laboratory, University of Wageningen. He is now the general director of the Plant Sciences Group, Wageningen UR, The Netherlands.

Ric C.H. De Vos studied Biology at the University of Amsterdam and obtained his PhD at the Free University of Amsterdam on plant physiology and toxicology. He is currently senior scientist in plant metabolomics and responsible for developing and implementing LC-MS-based metabolomics techniques at Plant Research International, Wageningen, The Netherlands.

Jacques Vervoort obtained a degree in Molecular Sciences and a PhD at the Biochemistry Laboratory, University of Wageningen, The Netherlands. He is an associate professor at the Biochemistry Laboratory and member of the Wageningen NMR Centre at University of Wageningen. He is interested in the development of MS and NMR technologies in biochemical proteomics and metabolomics applications.